ImmEx: IMMersive Text Documents EXploration System

Mario Cataldi
Università di Torino
Torino, Italy
cataldi@di.unito.it

Luigi Di Caro
Università di Torino
Torino, Italy
hardly dicaro@di.unito.it

Claudio Schifanella
Università di Torino
Torino, Italy
schi@di.unito.it

Abstract

Common search engines, especially web-based, rely on standard keyword-based queries and matching algorithms using word frequencies, topics recentness, documents authority and/or thesauri. However, even if those systems present efficient retrieval algorithms, they are not able to lead the user into an intuitive exploration of large data collections because of their cumbersome presentations of the results (e.g. large lists of entries). Moreover, these methods do not provide any mechanism to retrieve other relevant information associated to those contents and, even if query refinement methods are proposed, it is really hard to express it because of the user's inexperience and common lack of familiarity with terminology. Therefore, we propose ImmEx, a novel visual navigational system for an immersive exploration of text documents that overcomes these problems by leveraging the intuitiveness of semantically-related images, retrieved in real-time from popular image sharing services. ImmEx lets independently explore large text collection through a novel approach that exploits the directness of the images and their user-generated metadata. We finally analyze the efficiency and usability of the proposed system by providing case and user studies.

1. Introduction

The problem of indexing text collections is becoming more important than ever with the explosion of contents available on- and off-line. Considering these amounts of data, even if many retrieval systems still rely on user query models based on explicit keywords, there is a tendency to see browsing in contrast to directed searching, as a "don't-know-what-I-want" behaviour [1], or as a softer semi-directed or semi-structured searching process [4]. In fact, formulating queries represents a critical step in a retrieval process [7] because of the typical user's lack of familiarity with terminology, or because of terminology imprecisions. Moreover, the scope of a user query is often too broad and it can be very difficult to express with only few keywords; sometimes, users prefer to navigate within a topic rather than being dispatched to some system-relevant documents. Moreover, navigation by itself enables the users to understand the surrounding context. This behavior is generally called *orienteering* [14].

Based on these considerations, in this paper, we present a novel IMMersive EXploration system (*ImmEx*), for orienting the user within a large text collection by leveraging the intuitiveness of semantically-related images. The system dynamically indices text documents in real-time with respect to a set of semantically-related images, that represent the navigational interface of the system. Thus, the user is *immersed* in a visual space where it is possible to directly express abstract search interests by immediately selecting those figures that best represent the foci of interest. This visual approach permits to entrust the management of the query formulation problem to *ImmEx*, that formalizes the concepts expressed by the selected images exploiting the user-gererated metadata associated to them. The system relies on two steps:

- knowledge extraction: given a corpus of text contents, we formalize the semantics associated to the documents, to the terms (or keywords) in the documents and to the images which serve as guiding tools in the exploration process;
- exploration: given a set of retrieved images (downloaded in real-time from popular image sharing services) we define a novel mechanism to bridge the gap between the images and the text corpus that permits a more flexible navigation through the text collection.

At each iteration with *ImmEx*, the user can navigate within a set of proposed images, each associated to a set of related text documents. Each image can also act as a refinement for the previous user query to retrieve new, more focused, images (and consequently new text contents). Thus, in contrast with alternative text navigation methods, our approach directly leverages the expressiveness of the images to formalize queries over text corpora.

The paper is organized as follows: Sections 3 and 4 report the theoretical background of *ImmEx* while Section 5 explains the implemented system in details and provides a case study. In Section 6 we finally analyze the usability of the system based on user studies.

2. Related Works

There is a broad literature about ways of navigating collections of text documents. In [11], the authors present many reasons for using a navigational approach instead of document indexing, query matching and reformulation. Generally speaking, browsing operations need technologies aimed at comparing documents with user queries. An ordinary technique to do that was presented in [12], where both queries and documents are represented through vector space models. In our system, we make use of a slightly modified version [9], that assures a minumum value of relatedness for those keywords whose general frequency is low although they could be strongly relevant for a single topic.

As already discussed in the introduction, the problem of enriching a user query is an important task to bridge the gap between what she wants, and what the model actually needs to retrieve it. Query expansion methods try to solve the problem from different points of view based on linguistics [6], analysis of co-occurrences [17] [13], thesauri [16], relevance feedback [2], and others.

Another way of satisfying users' needs relies on the interaction with the system. Except for very simple cases, insight cannot be gained without automatic processing intertwined in the human analysis loop. The combination of automatic techniques and human analysis could offer new potentiality in presenting information in a way that can exploit human perceptual skills. Our system first leverages the human sensibility to the images, and then uses the associated users-tags to improve the navigation.

Images can generally give an at-a-glance insight about the meaning of something, therefore they can be used as semantic shortcuts in the user navigation path. Usually, they are treated as vectors of low-level features like color histograms, textures, image dimensions, and so on [5] [10].

In our system, instead of such image processing, we make use of associated metadata to guide the user in the navigation process. The exponential growth of user-generated contents on the web, thanks to sharing services like Flickr, represents in fact an easy way to exploit this concept by using user-generated tags. More in detail, once the user selects an image, the system leverages the associated metadata in combination with the current navigation history to further explore the collection of documents.

3. Knowledge Extraction

In order to be able to compare the text documents with the images, through their metadata information, a common model of knowledge representation is needed.

3.1. Document Definition

As in most information retrieval systems, the analysis process starts with the extraction of keywords (or terms) from the text collection. Precisely, given a corpus D with $n = |D|$ text documents, and a *corpus vocabulary* $V(D)$ with $v = |V(D)|$ keywords extracted from it [1], we preprocess each document $d \in D$ by associating to each of them a representative *document vector*, $\vec{dv}_d$, as

$$\vec{dv}_d = < w_{d,1}, w_{d,2}, ..., w_{d,v} >$$

which represents its semantics within the term space. In general, the $i-$th component $w_{d,i}$ of the vector $\vec{dv}_d$ represents the relatedness[2] of the $i-$th term with respect to the document d.

3.2. Keyword Definition

Conversely to the document definition, we model the semantics of the keywords according to their reciprocal correlations in the considered corpus D. In other words, we describe each keyword in terms of its contextual similarity with all the other keywords.

Let us consider the term "*bank*" in a set of news articles: depending on the focus expressed by the documents, it can refer to completely different concepts. In fact, from an economical point of view, its semantics can be currently defined by the association with other keywords like "finance", "crisis" and/or "money", while considering a set of medical articles, this term could be associated with keywords like "blood", "donation" or, for example, "transfusion". In other words, we can easily state that the meaning of each term can be differently interpretated depending on the context. We formalize this idea by associating to each keyword $k \in V(D)$ a *keyword vector*, $\vec{kv}_k$, formed by a set of weighted terms, that defines its context of use.

More formally, given a keyword k and its corresponding subset of documents D_k, we calculate (in a pre-processing phase) the correlation value $u_{k,z}$ with another keyword z relying on a probabilistic feedback mechanism [8]:

$$u_{k,z} = log \frac{\frac{r_{k,z}}{R_k - r_{k,z}}}{\frac{n_z - r_{k,z}}{N - n_z - R_k + r_{k,z}}} \times \left| \frac{r_{k,z}}{R_k} - \frac{n_z - r_{k,z}}{N - R_k} \right|,$$

where:

- N is the total number of documents;
- R_k is the number of documents containing k (it is equal to $|D_k|$); and

[1] Our system relies on *Wordnet* for morphological analysis like stemming and stopwords elimination.

[2] It is computed using the normalized term frequency presented in [9].

- n_z is the number of documents in the corpus containing the keyword z (it is equal to $|D_z|$);
- $r_{k,z}$ is the number of documents in D_k containing z.

We perform this operation for all keywords contained in at least one document. Notice that, given two keywords $k, z \in V(D)$, this feedback mechanism can return a negative correlation value $u_{k,z}$. In this case we set $u_{k,z} = 0$ in order to only consider those keywords that act positively in the contextual term defintion process.

Thus, given a term k, we obtain a *keyword vector*, $\vec{kv}_k$,

$$\vec{kv}_k =< u_{k,1}, u_{k,2}, \cdots, u_{k,v} >$$

which represents the contextual semantics of k in the corpus D provided by all the other weighted keywords in the considered vocabulary.

3.3. Image Definition

Since our aim is to help the user explore a set of text documents through a related image-space, we further need to formalize the semantics of this new type of content.

The general assumption of our work is that standard image processing techniques can be positively applied to evaluate the textures, the shapes, the colors, etc., of an image, but they are unable to provide valid interpretative mechanisms to decode the meaning they would like to express in the eye of the beholder. In order to do that, we used the information directly inserted by a user about her images, that provides a human-understandable coding of their semantics and summarizes the most important characteristics that the image should communicate to the users. Thus, we formalize the semantics of each image by exploiting the user-generated metadata associated to the images.

Considering the rapid increase of image sharing services on the web, millions of images are now publicly available through web portals like Flickr or Picasa Web. Moreover, many of them are indexed using user-generated metadata contents that help define their semantics. In particular, given an image i, we extract in real-time all the associated user-generated text information, including tags, description and title, in order to define its *image vocabulary* $V(i)$, where each keyword $k \in V(i)$ is also contained in the corpus vocabulary (i.e., $k \in V(i) \subset V(D)$). Notice that, for each image vocabulary IV_{img}, we only preserve the terms that appear in at least one text document in the considered corpus D, discarding those terms that are not defined in our context.

Then, using the contextual semantics of each $k \in V(i)$ provided by the related keyword vector $\vec{kv}_k$ (defined in Section 3.2) we create a unique *image vector* $\vec{iv}_i$ as

$$\vec{iv}_i =< t_{i,1}, t_{i,2}, \cdots, t_{i,v} >$$

where $v = |V(D)|$. Each element $t_{i,z}$ represents a keyword $z \in V(D)$ that helps define the semantics associated to the image i, and its weight is calculated as

$$t_{i,z} = \frac{1}{m} \times \sum_{k \in V(i)} u_{k,z}$$

where m is the cardinality of the image vocabulary ($m = |V(i)|$), k is a keyword in the considered image vocabulary $V(i)$ and $u_{k,z}$ represents the weight of the keyword z in the semantics of the keyword k expressed by $\vec{kv}_k[z]$.

This way, we compute in real-time an image vector with those user-generated metadata information (i.e., the tags) that are already contained in the context provided by the corpus.

4. Exploration and Visualization

In this Section we present the theoretical approach which serves as a basis for our image-based exploration of text corpora and the used visualization and interaction techniques.

4.1. Documents to Images Association

At this step, given an image i, its image vector $\vec{iv}_i$ provides a convenient way to identify the documents that best match with it. Semantic similarities between a given image i and a text document $d \in D$ can be computed in real-time by using their related vectors $\vec{iv}_i$ and $\vec{dv}_d$.

Thus, each image i is associated to a vector of documents

$$\vec{doc}_i =< s_{i,1}, s_{i,2}, \cdots, s_{i,n} >$$

where $n = |D|$ and the cosine similarity value $s_{i,d}$

$$s_{i,d} = cos(\vec{iv}_i, \vec{dv}_d),$$

quantifies how much the image i represents the knowledge expressed by the document d.

4.2. History Contextualization

The history of a navigation process is an important information to take into account when letting the user iteratively navigate through the contents. Our approach is easily adaptable to such scheme by leveraging the information from past queries. Thus, we revisit our initial definition of image vector $\vec{iv}_i$ with a new parameter α which manages the inclusion of the navigation history:

$$\vec{iv}_i =< \vec{iv}_{i,1} \times \alpha_{k_1}, \vec{iv}_{i,2} \times \alpha_{k_2}, \cdots, \vec{iv}_{i,m} \times \alpha_{k_m} >$$

where α_{k_z} represents the cosine similarity between the keyword vector $\vec{kv}_z$ and the user's current focus of interest defined by the previously visited image.

4.3. Displaying the Images

One of the primary goals of the proposed method is to ensure that the user can intuitively navigate into the document space through the visualized images. Therefore, given a set of images, we need to help the user understand the topics expressed by them by displaying the images in such a way that the similarities (or dissimilarities) are preserved.

For this purpose, we consider the set of documents associated to each image and we compute in real-time the cosine dissimilarity matrix M between their vectors:

$$M[i][j] = 1 - cos(\vec{doc}_i, \vec{doc}_j)$$

Then, we use a distance-preserving embedding technique (MDS [15]) to map these images onto a two-dimensional ordering, in order to be able to display them in a two-dimensional screen.

5. Experimental Setting

In order to assess the effectiveness of the proposed approach, we set up a real case scenario in which a hypothetical user would be able to:

- understand the context specified by the query;
- discover interpretations and subtopics of such domain;
- look for specific contents to read;
- refine the query based on the acquired knowledge.

5.1. Datasets

For this experimentation, we used the *AG* corpus of news articles[3] with more than 1 million of documents gathered from more than 2000 news sources. Each document has been associated by domain experts to a unique category. Thus, we use these categories to provide different navigational contexts to the user; in fact, she can choose among categories like "Entertainment", "Sci/Tech", "Business", "Sports", etc. This selection focuses the exploration process on the selected news context by limiting the considered documents to those associated to the selected category.

5.2. Case Study

In this Section, we present the interaction scheme with the system *ImmEx*[4], that implements the ideas reported in the paper.

In Figure 1 we show the interface of the proposed system: at the beginning of the exploration process, the user

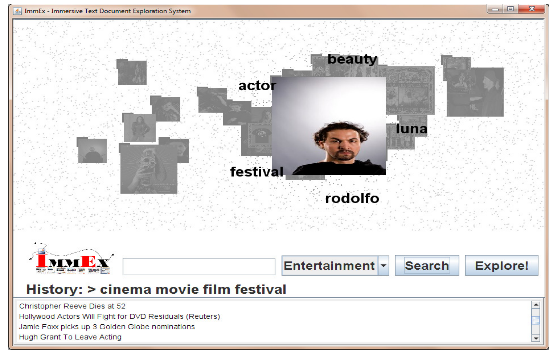

Figure 1. ***ImmEx*** **exploration system: the visualized image, when clicked, allows the user to easily specialize her exploration by focusing on concepts like actor, or beauty, in the "Entertainment" context.**

starts the navigation by selecting the preferred category that defines her exploration context. (in the presented example, the user selects "Entertainment"). Then, she can start the navigation by typing a keyword query (in the presented case "cinema") or, considering the initial orienteering idea, by letting the system select an exploration route (by clicking on "Explore!", the system randomly chooses a keyword in the selected context).

At this point, *ImmEx* analyzes the semantics associated to the term (computed during the pre-processing phase, Section 3.2) and it contextualizes the query by associating, to the original term, other relevant keywords that help define the concept expressed by the user ("film", "festival" and "movie" in the example in Figure 1). *ImmEx* leverages this information to retrieve, in real-time, correlated images (by formalizing a query to Picasa Web Service that combines the initial term with the retrieved semantically related keywords). The resulting images are then inspected in order to extract the user-generated metadata terms; thus, the system computes the image vector for each of them by leveraging the contextual definition of each metadata term in the considered context (Section 3.3) and it associates to each image a set of related text documents (Section 4.1). It finally places the images in the screen according to their mutual dissimilaries in terms of associated documents (Section 4.3). The exploration within the document space is then guided by the system through simple visualization metaphors: the size of each image can help understand the number of the associated documents (the bigger the size, the higher the number of associated documents), while its position on the screen permits to easily identify semantic correlations among the retrieved images (the closer they are, the more similar the associated documents).

By simply moving the mouse over an image, it is then possible to carefully analyze it (*ImmEx* enlarges it and fades out the other images); moreover, in order to understand

[3]http://www.di.unipi.it/∼gulli/AG_corpus_of_news_articles.html
[4]Sample video at http://www.di.unito.it/∼schi/immex/immex.mov

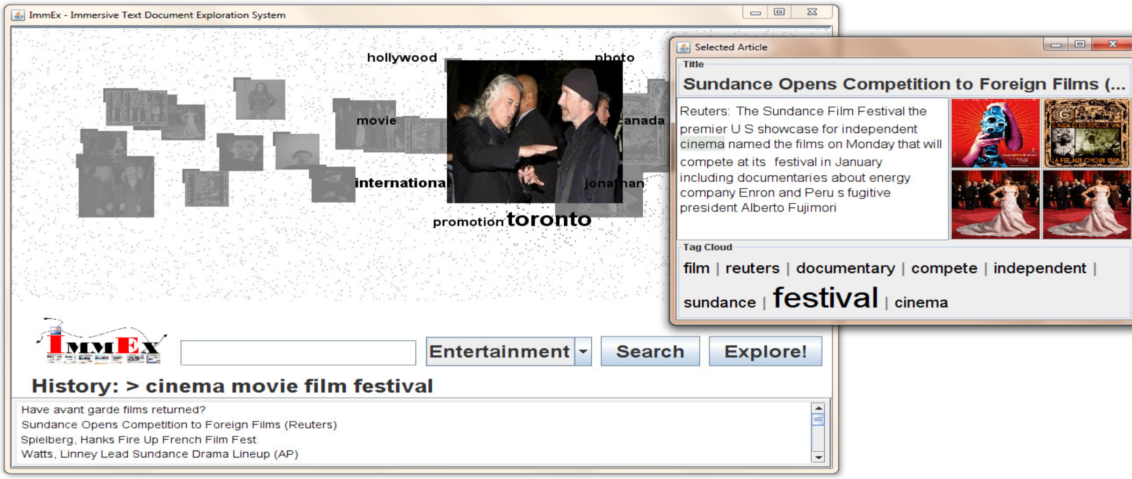

Figure 2. The user analyzes an image about a gala at Toronto Film Festival (and the related metadata terms). By opening a related text document, *ImmEx* reports the entire original text, the related tag cloud and a minimal set of representative images.

the associated semantics, the most relevant user-generated metadata keywords (in terms of frequency in the related metadata) are shown. In Figure 2, the user visualizes an image that refers to a gala inauguration at the Toronto Film Festival. The metadata information reports terms like "festival", "international" and "hollywood"; thus, the user can easily understand that this image represents a subtopic (of the considered concept "cinema") related to public events like "cinema competition". In constrast, in Figure 1 the represented image defines another subtopic related to the movie stars (i.e. the reported metadata terms highlight the concept by presenting terms like "actor" and "beauty"). Therefore, using the image visualized by *ImmEx*, the user entrusts the management of the query formulation to the system that formalizes the abstract concepts expressed by the selected image by using its user-gererated metadata.

Besides, if a user left-clicks on the image, an ordered list of related news articles is presented (showing the title) at the bottom of the interface. By clicking on one of them (Figure 2), *ImmEx* shows the entire text article (where the metadata keywords that guided the association with the selected image are highlighted) with some other functional information; in particular, *ImmEx* provides an article's tag cloud (that highlights the most relevant terms of the articles in the selected context category), and a minimal set of semantically correlated images (by measuring the cosine similarity among its document vector and the image vectors of the retrieved ones).

If the user is interested in focusing on a subtopic expressed by a single image, she just needs to double-click on it to trigger a refinement of the current search interest (Section 4.2). The history of the navigation process is preserved and shown as a list of keywords (separated by arrows) at the bottom of the interface. The user can always clear the history (restarting the exploration process) by typing a new keyword query or by using the orienteering approach through the button "Explore!" (that restarts the entire process with a randomly selected keyword).

6. User Study

In order to analyze the benefits of using the proposed IMMersive text documents EXploration system, we also conducted a user study (similarly to [3]) and evaluated the feedback of 14 users when exploring the previously described *AG* corpus of news articles. The users are representative of various age ranges, backgrounds, jobs and education levels, and have intermediate web ability (they are not computer scientists or domain experts).

We analyzed the users feedback about the accuracy of the proposed documents-to-image association approach (Section 4.3) and we also studied their opinions about the usability of the *ImmEx* exploration system.

6.1. Image Association Accuracy

Given three randomly selected news articles, extracted from the considered AG corpus of documents, we asked the users to evaluate their associations with images (extracted in real-time by *ImmEx* from Picasa Web Service) performed through the documents-to-images association step (Section 4.3). In particular, given the three selected articles, we asked the users to read the original text articles and analyze the images to which they were associated; thus, we asked to express their opinions about the coherence of these associations by using a 5-point scale rating (where 1 expresses an inadequate association and 5 a very semantically-coherent one). The results are shown in Table 1.

The results indicate that, for all the proposed news articles (different in size and topics), the users agreed about the

Context: AG Corpus	
	avg. users rating
1st document	4.42
2nd document	4.57
3th document	4.28
total avg.	4.42

Table 1. User Study: average users ratings (expressed by using a 5-point scale) about the association with images regarding three randomly selected news articles.

efficacy of our image association approach by judging as very coherent the correlation with the proposed images (average rating of 4.42). Moreover, it is possible to note that the behaviour of this association process is really consistent, while it does not present significant differences among the considered cases (the feedback of the users is included between 4.28 and 4.57) even if their main topics differ considerably. Therefore, this documents-to-images association process is independent from the topics of the articles and it is able to retrieve the most suitable images from web share services by positively leveraging the user-generated information.

6.2. Subjective Questionnaire Measures

After the study, each user also completed a brief questionnaire which included three questions ("Is the proposed exploration system easy to use?", "Is the proposed exploration system pleasant to use?" and "Does *ImmEx* help the user explore the text corpus?"); again, the users could quantify the responses using a 5-point scale.

The users reported that *ImmEx* was really "easy to use" (avarage rating of 4.14) and provides a very agreeable navigation experience into the considered text documents corpus (4.71). Therefore, the main assumption of our approach was confirmed by the users' feedback while they reported that the intuitiveness of the images can be positively leveraged to improve the exploration of text corpora and their semantics can be easily understood by common web users. In fact, as originally supposed, in a world saturated by continuous visual-audio inputs, the users are generally used to manage and decipher the meanings of visual messages and thus this methapor can be positively used for text exploration purposes.

Moreover, the users agreed to declare that *ImmEx* provides a convenient way to explore the considered text documents (average rating of 4.28). Finally, the users confirmed that the hard task of the query formulation (and refinement) can be succesfully entrusted to images that can help bridge the gap between users' abstract search desires and the query formulation problem.

7. Conclusions

In this paper, we presented *ImmEx*, a novel visual navigational system where the user is *immersed* in a image-based space in which it is possible to directly express abstract search interests by immediately selecting those images that best represent the user's foci of interest. We studied the proposed approach by analyzing the features of *ImmEx*, which implements the idea reported in this paper, and we evaluated the users feedback by providing user studies.

References

[1] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 1989.

[2] M. Cataldi, C. Schifanella, K. Candan, M. Sapino, and L. Di Caro. CoSeNa: a context-based search and navigation system. In *MEDES*. ACM, 2009.

[3] H. Chen and S. Dumais. Bringing order to the web: automatically categorizing search results. In *CHI*. ACM, 2000.

[4] D. Ellis. A behavioral approach to information retrieval system design. *J. Doc.*, 45(3):171–212, 1989.

[5] C. Frankel, M. Swain, and V. Athitsos. Webseer: An image search engine for the world wide web. 1996.

[6] G. Grefenstette. Short query linguistic expansion techniques: Palliating one-word queries by providing intermediate structure to text. *Information Extraction*, 1997.

[7] S. Y. Rieh and H. Xie. Analysis of multiple query reformulations on the web: the interactive information retrieval context. *Inf. Process. Manage.*, 2006.

[8] I. Ruthven and M. Lalmas. A survey on the use of relevance feedback for information access systems. *Knowl. Eng. Rev.*, 2003.

[9] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. In *Information Proc. and Man.*, 1988.

[10] S. Sclaroff, L. Taycher, and M. La Cascia. Imagerover: A content-based image browser for the world wide web. In *Content-Based Access of Image and Video Libraries*. IEEE, 2002.

[11] S. P. Shashank. Navigation-aided retrieval. In *Proc of WWW'07*, pages 391–400.

[12] A. Singhal and G. Salton. Automatic text browsing using vector space model. In *Proceedings of the Dual-Use Technologies and Applications Conference*, 1995.

[13] K. Sparck Jones. Notes and references on early automatic classification work. *SIGIR Forum*, 25(1):10–17, 1991.

[14] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *CHI*. ACM, 2004.

[15] W. S. Torgerson. *Theory and methods of scaling*. R.E. Krieger Pub. Co.

[16] E. Voorhees. Query expansion using lexical-semantic relations. In *SIGIR*, 1994.

[17] J. Xu and W. B. Croft. Improving the effectiveness of information retrieval with local context analysis. *Inf. Syst.*, 2000.